Fatih Porikli, Shiguang Shan, Cees Snoek,
Rahul Sukthankar, and Xiaogang Wang

Deep Learning for Visual Understanding: Part 2

Visual perception is one of our most essential and fundamental abilities that enables us to make sense of what our eyes see and interpret the world that surrounds us. It allows us to function and, thus, our civilization to survive. No sensory loss is more debilitating than blindness as we are, above all, visual beings. Close your eyes for a moment after reading this sentence and try grabbing something in front of you, navigating your way in your environment, or just walking straight, reading a book, playing a game, or perhaps learning something new. Of course, please do not attempt to drive a vehicle. As you would realize again and appreciate profoundly, we owe so much to this amazing facility. It is no coincidence that most of the electrical activity in the human brain and most of its cerebral cortex is associated with visual understanding.

Computer vision is the field of study that develops solutions for visual perception. In other words, it aims to make computers understand the seen data in the same way that human vision does. It incorporates several scientific disciplines such as signal processing, machine learning, applied mathematics, sensing, geometry, optimization, statistics, and data sciences to name a few. It is concerned with the extraction, modeling, analysis, and use of information from a single image or a sequence of images across a spectrum of modalities for building intelligent systems.

As our visual perception of the world is reflected in our ability to make decisions through what we see, providing such analytical capabilities to computers makes it possible to design remarkable applications that enhance our lives. Computer vision solutions are acting everywhere, including in our

Visual perception is one of our most essential and fundamental abilities that enables us to make sense of what our eyes see and interpret the world that surrounds us.

- computer mouse, determining its motion
- phones, reading our fingerprints
- cameras, controlling lenses
- mail centers, sorting parcels
- warehouse robots, retrieving packages
- gateways, recognizing faces
- vehicles, assisting drivers
- hospitals, diagnosing medical problems
- factories, performing inspections
- farmlands, harvesting produce
- dressers, checking the style of our outfits.

As well as revolutionizing technologies for autonomous vehicles and virtual reality devices, it will soon unfold a transformative and disruptive impact on our culture and economy.

On the journey of developing algorithms that can match human visual perception, most of the progress happened within the last decade with the rebirth of artificial neural networks in computer vision, in particular, convolutional architectures. Ascribing to their complex and layered structures, a broader family of data-driven machine-learning methods based on neural network models today is called *deep learning*. An illustration of common deep-learning networks such as convolutional neural networks, autoencoders, and generative-adversarial networks (GANs) can be seen in Figure 1, and a very comprehensive discussion of different deep-learning techniques for visual understanding also can be found in the tutorial articles in the first part of this special issue in the November 2017 issue of *IEEE Signal Processing Magazine* (*SPM*).

There are many compelling advantages of deep-learning methods. In their cascaded layers that can contain hundreds of millions of parameters, they can model highly nonlinear functions. With their pooling layers that can generate multiple levels of representations corresponding to different levels of abstraction, they can coalesce the information from local and global receptive fields. They can run efficiently on parallel processors with their feed-forward characteristics. Since they learn what part of the data is relevant and discriminative from training samples automatically, they are not limited to handcrafted descriptors and manually defined

Digital Object Identifier 10.1109/MSP.2017.2766286
Date of publication: 9 January 2018

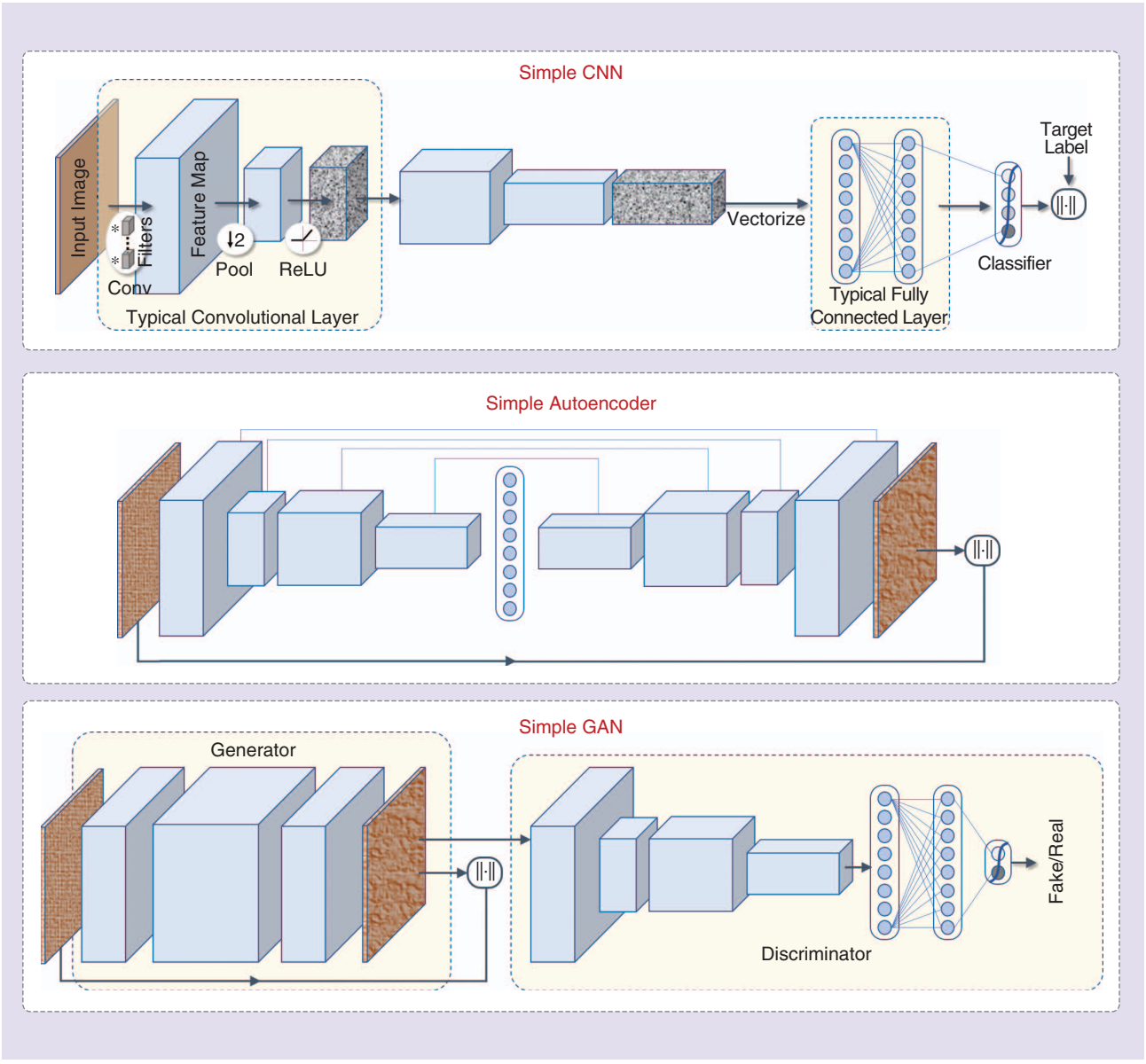

FIGURE 1. Simple deep-learning architectures. (Figure courtesy of Fatih Porikli.)

transformations. Most importantly, they can learn from their mistakes when provided with such cases and become superior as the amount of training data increases. The success of deep-learning methods also reflects on the volume of the scientific publications. Deep-learning-related articles in main computer vision venues boosted from fewer than 100 in 2012 to an astounding level of more than 1,000 in 2017.

This edition also has three articles on popular areas of GANs, deep regression Bayesian networks, and model compression and acceleration.

The November 2017 special issue of *SPM* on deep learning for visual understanding surveyed deep-learning solutions under reinforcement; weakly supervised and multimodal settings, investigated their robustness; and presented overviews of their applications in domain adaptation, hashing, semantic segmentation, metric learning, inverse problems in imaging, image-to-text generation, and picture-quality assessment.

Complementing these topics in this second part of the special issue on deep learning for visual understanding, we continue providing tutorials on deep-learning techniques for understanding face images, salient and category-specific object detection, superresolution, denoising, deblurring, compressive sensing, zero-shot recognition, and conditional random fields. This edition also has three articles on popular areas of GANs, deep regression Bayesian networks, and model compression and acceleration. We hope these tutorial articles will foster further discussions and facilitate

the application of deep-learning techniques for computer vision to the other areas of signal processing. Once again, we welcome you to explore all of these articles as well as the amazing field of deep learning, and we wish you a wonderful new year.

Acknowledgments

We thank all of the contributors for their outstanding articles that individually and as a whole, addressed relevant and timely aspects in deep learning, highlighting the key roles that they take in computer vision. We are grateful to *SPM*'s Editor-in-Chief Prof. Min Wu and Managing Editor Jessica Welsh for their continuous support and assistance. We very much enjoyed putting together this special issue, and we do believe that our readers will enjoy it twice as much.

Meet the guest editors

Fatih Porikli (fatih.porikli@anu.edu.au) received his B.Sc. degree in electrical engineering from Bilkent University, Turkey, in 1992 and his Ph.D. degree in electrical and computer engineering from New York University in 2002. He is an IEEE Fellow and a professor at Australian National University. He is also the chief scientist at Huawei, Santa Clara, California. Previously, he served as the Computer Vision Research group leader at National ICT Australia and distinguished scientist at Mitsubishi Electric Research Laboratories. His research interests include computer vision and machine learning with commercial applications in autonomous vehicles, video surveillance, visual inspection, robotics, and medical systems. He received the R&D100 Scientist of the Year Award in 2006, won five Best Paper Awards at IEEE conferences, and invented 71 patents.

Shiguang Shan (sgshan@ict.ac.cn) received his B.S.E. and M.S.E. degrees in computer science from Harbin Institute of Technology, China, in 1997 and 1999, respectively. He received his Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing, in 2004, where he has been a full professor since 2010 and is now the deputy director of the CAS Key Lab of Intelligent Information Processing. His research interests include computer vision, pattern recognition, and machine learning. He has published more than 200 papers in these areas. He served as area chair for many international conferences and is an associate editor of several journals, including *IEEE Transactions on Image Processing*, *Computer Vision and Image Understanding*, *Neurocomputing*, and *Pattern Recognition Letters*.

Cees Snoek (cgmsnoek@uva.nl) received the M.Sc. degree in business information systems in 2000 and the Ph.D. degree in computer science in 2005, both from the University of Amsterdam, The Netherlands. He is currently a director of the QUVA Lab, the joint research lab of Qualcomm and the University of Amsterdam, on deep learning and computer vision. He is also a principal engineer/manager at Qualcomm and an associate professor at the University of Amsterdam. His research interests focus on video and image recognition. He has published more than 200 refereed book chapters and journal and conference papers. He received a Veni Talent Award, a Fulbright Junior Scholarship, a Vidi Talent Award, and The Netherlands Prize for Computer Science Research, all for research excellence.

Rahul Sukthankar (rahulsukthankar@gmail.com) received his B.S.E. degree in computer science from Princeton University, New Jersey, in 1991 and his Ph.D. degree in robotics from Carnegie Mellon, Pittsburgh, Pennsylvania, in 1997. He leads research efforts in computer vision, machine learning, and robotics at Google. He is also an adjunct research professor with the Robotics Institute at Carnegie Mellon and courtesy faculty at the University of Central Florida. Previously, he was a senior principal researcher at Intel Labs, a senior researcher at HP/Compaq Labs, and a research scientist at Just Research. He has organized several workshops and conferences and currently serves as the editor-in-chief of *Machine Vision and Applications*.

Xiaogang Wang (xgwang@ee.cuhk.edu.hk) received his bachelor's degree in electronic engineering and information science from the Special Class of Gifted Young at the University of Science and Technology of China in 2001, his M.Phil. degree in information engineering from the Chinese University of Hong Kong in 2004, and his Ph.D. degree in computer science from the Massachusetts Institute of Technology in 2009. He has been an associate professor in the Department of Electronic Engineering at the Chinese University of Hong Kong since August 2009. He received the PAMI Young Research Award Honorable Mention in 2016. He is an associate editor of *Image and Visual Computing Journal*, *Computer Vision and Image Understanding*, and *IEEE Transactions on Circuit Systems and Video Technology*.

SP